Evaluation plan for the Dutch Structural Business Statistics questionnaires: using output to guide input improvements

Giesen, Deirdre

EVALUATION PLAN FOR THE DUTCH STRUCTURAL BUSINESS STATISTICS QUESTIONNAIRES: USING OUTPUT TO GUIDE INPUT IMPROVEMENTS

DEIRDRE GIESEN[1]

1. Introduction

In establishment surveys issues of usability and respondent-friendliness are often neglected. Dillman (2000, p 345) tellingly describes the implicit model for government business surveys as "a Cost Compensation Model". In this model the goal to minimize monetary cost determines the questionnaire design and implementation practices. To compensate for the resulting flaws in the design of the instruments the data collection agency relies on the fact that the participation in most government establishment surveys is mandatory. However, recently there has been more and more interest in the improvement of data collection for establishment surveys to reduce response burden and increase data quality (e.g. Goldenberg et al. 2002, Jones 2003, Hak & Willimack 2003).

A similar trend can be seen at Statistics Netherlands (SN). In a time of decreasing resources, the efficiency of the production of statistical information is of utmost concern. To work more efficiently Statistics Netherlands has redesigned the statistical process for the Structural Business Statistics (SBS). In 1998 a project was started with the aim to integrate and standardize the data collection, the editing, and the publication of the SBS. This project is now known as IMPECT (IMPlementation of the EConomical Transformation process). In the first years of IMPECT emphasis was put on creating the logistics of the system. Recently, the attention has moved to the evaluation and improvement of the content of the questionnaires. For 2004 an evaluation and revision of

1 The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

the SBS questionnaires is planned. This paper describes the strategy we developed to evaluate this set of almost 200 questionnaires which vary depending on branch and size of the establishments surveyed.

2. Goal of the evaluation

The goal of the evaluation is to improve the phrasing of the questions and the lay out of the questionnaires in order to increase the quality of the data and reduce the response burden[2]. These goals may prove to be conflicting in many situations, as better measurement at the micro level often means more detailed questioning.

3. The Structural Business Statistics Questionnaires

The SBS questionnaires, also known as the 'Production Surveys', measure a number of indicators of the activity and performance of enterprises in manufacturing, construction, trade, transport, commercial services, energy and water. Variables collected include detailed information on turnover and expenditure of the past year. Important sources for the asked information are the business balance sheets and profit-and-loss accounts. The specification of items such as personnel costs and housing or stock value, however, calls for a consultation of other administrative records.

The data is collected by mail-out, mail-back forms. Each questionnaire is sent out with a so called 'remark sheet' which respondents should use to make comments about the questionnaires, make changes in the name or address of the firm or to ask for delay of the deadline for sending back the questionnaire. Responding to the SBS questionnaires is mandatory. Of all SN establishment surveys, the SBS rank second with respect to response burden, measured as the time needed to fill out the questionnaire. Questionnaires of more than 15 pages are typical. All size classes are covered, but for smaller firms sampling is used. In 2002 almost 80,000 questionnaires were sent out, with a response rate of 68%.

One of the goals of IMPECT was to uniformize the SBS questionnaires in order to gain more efficiency in the data collection process. The process of data collection has indeed been standardized completely. All questionnaires are automatically generated from a system called LogiQuest. The questionnaires are uniformized to a large extent at the level of the *variables* measured. However, similar variables can (and must) be measured with

2 Other projects at SN explore different ways to reduce response burden, for example the possibilities of automatic tapping of records and reducing the number of questions.

different questions or response categories, since it would be rather difficult to specify the turnover of a shoe store and a construction company with the same items. Thus, all SBS questionnaires have a uniform part that is the same for all branches, and a part with branch specific questions. Within each branch there is also a short and a longer form, depending on the size of the businesses (or more correctly: the level of detail needed to construct the statistics). The combination of size and branch specific *questions* results in 183 different questionnaires. Within these 183 groups of questionnaires forms are not always identical, as they may contain product lists that are uniquely compiled for a specific establishment, according to information already available from that firm.

4. Strategy for Testing

We have chosen a strategy that mixes both quantitative and qualitative methods and where the cheaper methods are used to prioritize the more expensive qualitative methods. Given the large amount of different SBS questionnaires and the heterogeneity of respondents it is simply not feasible to test all different questionnaires qualitatively. Fortunately, as the questionnaires have already been in the field, there is quite some process and survey data available that we can use in our evaluation. This triangulation is comparable to the formative evaluation described by the UK Office of National Statistics (ONS) as part of their framework of reviewing data collection instruments in business surveys (Jones, 2003).

4.1 Office based analysis of the questionnaires

Our first step in the evaluation is an office based analysis of the questionnaires. The SBS questionnaires have been in the field for three years now. This means that there are survey data and process information available that can be used to make inferences about the questionnaire. We will also use qualitative data from different sources available at SN, such as the data editors who work with the questionnaire. The goal is to reach an empirically grounded overall analysis of the SBS questionnaires. After this round we should know which questionnaires, questions and types of respondents are most problematic with respect to data quality and response burden.

Survey data and process data

Differences in *unit response* may be an indicator of problematic questionnaires. We must therefore take into consideration how branch and size class characteristics relate to unit response for the SBS questionnaires. If there are groups with particularly low or late response rates, it will be useful to further investigate whether these differences can be attributed to characteristics of the questionnaires. For example, one could test the relation

between the number of specific questions in a questionnaire and the likelihood of a timely response. Patterns in *data quality* may be another important indicator of the quality of the questionnaire. We will examine three ways to operationalize the data quality at the level of questions: 1) items non-response, 2) plausibility of the data as calculated in the editing process 3) percentage of changes made in the data during the editing process. If we manage to develop useful quality measures at the item level, these data present excellent material to investigate the effect of questionnaire and respondent characteristics on data quality.

Content analysis of respondents' remarks and filled out questionnaires

All remarks made by respondents about the questionnaire are documented in *LogiQuest*. This system contains both the information provided on the so called 'remark sheet' (see paragraph 3) as well as remarks about the questionnaire that are made to the call center staff. So far we have not systematically looked into that data base and it will be interesting to see if a content analysis of these remarks will provide useful information for the evaluation of questionnaires.

At ONS samples of questionnaire images are analyzed as part of the process of questionnaire evaluation (Jones, Williams & Thomas, 2003). A first look at some filled out SBS questionnaires shows that crossed out questions, accolades written in the margin to group specified items and comments about the questions give interesting insight in parts of the response process. A very attractive feature of this analysis is that it can be done systematically.

Interviews with SN staff

In their work with respondents, questionnaires and the collected data, employees from different departments of SN have gained insights into possible strength and weaknesses of the questionnaires. We will organize a round of focus groups and open interviews to make these ideas and information available to our evaluation. Four types of informants can be distinguished:

Interviews with our *field officers* - who visit non-responding firms and sometimes help firms to fill out questionnaires - have proved very useful in previous projects (Snijkers, 2000; Giesen, 2003). Rowlands, Eldrigde and Williams (2002) found that *data editors* also provide important and new insights to questionnaire problems. A third group of relevant informants are the *call center staff* who make the non-response follow up calls and answer the first helpdesk request by respondents. A last, but not least important group of SN employees to talk to are the *users of the data*, the people working on the analyses and publication of the data. They may know of patterns in the data that indicate

problematic questions. Experiences with household surveys show that the data users can provide important points of interest for testing and evaluation. Also, interviews with data users on their ideas of possible flaws in the questionnaires present a great opportunity to involve them in the evaluation of 'their' questionnaires. This will hopefully help create commitment among this group for any changes in the questionnaire made later on.

Expert Review

Next to the analyses of the existing information about how the questionnaires work in the field, we will give a small sample of typical SBS questionnaires to experts in the field of questionnaire design. If possible, we will present the questionnaires together with the results of the review described above and a first concept of the field-tests planned. The experts will be asked to comment on the questionnaires, our conclusions and plans so far, as well as to come up with possible solutions for problems already discovered.

4.2 Diagnoses of questionnaire problems in the field

The round of office based analyses should provide us with a good overview of the most problematic questions and questionnaires and the groups of respondents where these problems occur most. Some of these problems may be straightforward and it will be easy to decide if and how they can be solved. In other cases we will need information from respondents to analyze why questions do not work for them and how we can improve these. We will use the results of the office based analyses to decide where we will focus our fieldwork. When prioritizing our limited capacity for field testing, we will also consider practicalities such as the importance of problematic questions or groups of respondents for the output of the survey and whether or not a question can be changed.

The goal of this second step is to diagnose the problems found and look for possible remedies. Recent experiences at SN have given us a good idea on how we can collect useful information on problems with establishment questionnaires by studying the response process in the field. A pilot at SN by Hak and van Sebille (2002) has shown that focused on-site interviews yield useful insights in problems of the SBS questionnaires. For this pilot four constructing companies that were known as good respondents were visited. The goal of the focused interviews was to approach an observation of the actual process of filling out the questionnaire as close as possible. With the already completed and returned questionnaire at hand, the researcher and field officer reconstructed the response process with the respondent. This meant that item for item it was assessed whether and how respondents had come to an answer. Respondents were able and willing to explain if and how they had estimated or calculated the numbers given. This detailed

information revealed misinterpretations of questions and definitions and satisficing behavior.

One of the conclusions of this pilot was that on-site observation of the actual response process might very well be possible. This was successfully tried in the field in the context of the evaluation of the Transportation Survey (Giesen, 2003). Here a methodologist, a field officer and a camera man visited three respondents. The respondents were observed and filmed while filling out an electronic questionnaire.

During these visits we had three main goals: observing what respondents do, understanding why they do it and collecting good survey data. Firstly, we wanted to observe how respondents go about when they work with the questionnaire. For this purpose we encouraged the respondents to start with the questionnaire as they would if we had not been present. During this phase we tried to restrict the interaction with respondent to questions that were necessary to clarify what the respondent was doing at the time ("What are you looking for now?" or "What are those records?"). Secondly, we needed insight in *why* respondents filled out the questions the way they did and how they evaluated the instrument. For this purpose, after the completion of the questionnaire, we asked the respondents how they had understood and answered crucial questions and how they felt about the user-friendliness of the instrument. Thirdly, as real data were collected in these sessions and respondents were likely to have to complete similar questionnaires in the future, we wanted to correct errors respondents had made and to explain how they should have done it.

It proved rather difficult to strictly distinguish the three phases and goals of the visit. Especially when respondents got stuck in the questionnaire it was sometimes impossible not to intervene eventually. Without a doubt, our mere presence and our interventions will have influenced the motivation of the respondents and the ease of filling out the questionnaire. However, even with this bias, we gained insight in where and how respondents made errors in the questionnaire and what aspects of the questionnaire were particularly burdensome. We believe some of the problems found, especially with respect to navigational issues, could only have been obtained by some form of on-site behavioral observation.

Our experiences so far indicate that on-site observation or – if actual observation is not possible – retrospective focused interviews on-site, yield rich and useful data to evaluate and improve questionnaires. It is needless to say that these rich data come at a high cost. It may take several days to organize and actually do an on-site observation.

4.3 Development and testing of improved questionnaire

A round of qualitative fieldwork with on-site observation should result in recommendations for the improvement of the SBS questionnaires. These changes in the questionnaire have to be tested with respondents, to make sure that the changes are in fact improvements for the problems found and have not created new problems. The scope and methods for this test round will be developed when we have the results from the first two steps. Than we will know how many changes have been made to the questionnaires and we can assess the risks of these changes.

5. Future plans

The implementation and further development of the testing strategy will undoubtedly give us a lot of information about the usefulness of our evaluation methods. Besides the practical goal to improve the SBS questionnaires, we also have a research goal to increase our insights in the response process of establishments and the best ways to study these processes. In this research project questions will be addressed such as: Can we indeed distinguish problematic questions by desk research? Which of the sources of information available about questionnaires in the field are more useful to evaluate questionnaires and which are less useful? Can and should we incorporate these kinds of evaluation in a process of systematic review?

Literature

Dillman, D. A., 2000: Procedures for conducting government-sponsored establishment surveys: comparisons of the total design method (TDM), a traditional cost-compensations model, and tailored design. In ICES II, Proceedings of the Second International Conference on Establishment Surveys, Alexandria, VA: American Statistical Association, pp. 343-352.

Giesen, D., 2003: Het gebruiksgemak van de elektronische vragenlijst Verkeer en Vervoer 2003. [The user-friendliness of the electronic transportation questionnaire 2003] Heerlen: Statistics Netherlands.

Goldenberg, K. L./Anderson, A. E./Willimack, D.K./Freedman, S.R./Rutchnick, R.H./Moy, L.M., 2002: Experiences Implementing Establishment Survey Questionnaire Development and Testing at Selected U.S. Government Agencies. Paper presented at QDET conference, June 2002, Charleston, SC.

Hak, T./van Sebille, M., 2002: Het respons proces bij bedrijfsenquêtes. Verslag van een pilot studie. [The response process in establishment surveys. Report of a pilot study] Rotterdam/Voorburg: Erasmus Research Institute of Management/Statistics Netherlands.

Jones, J., 2003: A framework for evaluating and redesigning data collection instruments. Survey Methodology Bulletin. Office for Nationals Statistics, 61:4-9

Jones, J./Williams, S./Thomas, M., 2003: The Use of Administrative Sources in the Evaluation & Design of Data Collection Instruments. Paper presented at the QUEST Workshop, October 21-23 2003, Mannheim, Germany.

Jong, A. de, 2002: Uni-edit: Standardized processing of structural business statistics in the Netherlands. Paper presented at Conference of European Statisticians, UNECE Work Session on Statistical Data Editing, May 2002, Helsinki, Finland.

Rowlands, O./Eldridge, J/Williams, S., 2002: Expert review followed by interviews with editing staff – effective first steps in the testing process for business surveys. Paper presented at QDET conference, June 2002, Charleston, SC.

Snijkers, G., 2000: Eindverslag focusgroepen met nieuw concept omslagvel PS (IMPECT). [Final report focus groups on new concept cover Production Survey IMPECT] Heerlen: Statistics Netherlands.

Contact

Deirdre Giesen
Statistics Netherlands
Kloosterweg 1
NL-6412 CN Heerlen
The Netherlands
email: igin@cbs.nl